

# Toward physical realizations of thermodynamic resource theories<sup>a</sup>

Nicole Yunger Halpern<sup>b1</sup>

<sup>1</sup>*Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA*

(Dated: January 19, 2017)

Conventional statistical mechanics describes large systems and averages over many particles or over many trials. But work, heat, and entropy impact the small scales that experimentalists can increasingly control, e.g., in single-molecule experiments. The statistical mechanics of small scales has been quantified with two toolkits developed in quantum information theory: *resource theories* and *one-shot information theory*. The field has boomed recently, but the theorems amassed have hardly impacted experiments. Can thermodynamic resource theories be realized experimentally? Via what steps can we shift the theory toward physical realizations? Should we care? I present eleven opportunities in physically realizing thermodynamic resource theories.

PACS numbers: 05.70.Ln, 03.67.-a 89.70.Cf, 05.70.-a,

## CONTENTS

I. Introduction	1
II. Technical introduction	2
A. Resource theories	2
B. Thermodynamic resource theories	3
C. One-shot statistical mechanics	3
III. Opportunities in physically realizing thermodynamic resource theories	5
A. What merits realization? How, in principle, can we realize it?	5
1. What would epitomize realizations of the one-shot statistical mechanics developed with TRTs?	5
2. How can we test predictions of maximal efficiencies?	7
3. Which operations should be performed to test TRT results?	8
B. Enhancing TRTs' realism	9
1. Should free operations reflect experimentalists' capabilities better? If so, how?	9
2. Can nonconservation of energy model systems' openness and model the impreciseness with which unitaries are implemented?	11
3. How should TRTs model measurements?	11
4. Can a measure of $\epsilon$ -closeness lead to testable predictions and away from embezzlement?	12
5. Can a definition of $\epsilon$ -smoothing more naturally model errors and failure?	14
6. How should TRTs model continuous spectra?	15
C. More-out-of-the-way opportunities	15
1. Can TRTs model quantum field theories?	16
2. Can fluctuation theorems bridge TRTs to experiments?	16
IV. Conclusions	17
Acknowledgements	17
References	18

## I. INTRODUCTION

“This is your arch-nemesis.”

<sup>a</sup>To appear in the conference proceedings for “Information and Interaction: Eddington, Wheeler, and the Limits of Knowledge,” in the FQXi subset of Springer’s Frontiers book collection.

<sup>b</sup>E-mail: nicoleyh@caltech.edu

The thank-you slide of my presentation remained onscreen, and the question-and-answer session had begun. I was presenting a seminar about *thermodynamic resource theories* (TRTs), models developed by quantum-information theorists for small-scale exchanges of heat and work [1, 2]. The audience consisted of condensed-matter physicists who studied graphene and photonic crystals. I was beginning to regret my topic’s abstractness.

The question-asker pointed at a listener.

“This is an experimentalist,” he continued, “your arch-nemesis. What implications does your theory have for his lab? Does it have any? Why should he care?”

I could have answered better. I apologized that quantum-information theorists, reared on the rarefied air of Dirac bras and kets, had developed TRTs. I recalled the baby steps with which science sometimes migrates from theory to experiment. I could have advocated for bounding, with idealizations, efficiencies achievable in labs. I should have invoked the connections being developed with fluctuation relations [3–5], statistical mechanical theorems that have withstood experimental tests [6–13].

The crowd looked unconvinced, but I scored one point: the experimentalist was not my arch-nemesis.

“My new friend,” I corrected the questioner.

His question has burned in my mind for two years. Experiments have inspired, but not guided, TRTs. TRTs have yet to drive experiments. Can we strengthen the connection between TRTs and the natural world? If so, what tools must resource theorists develop to predict outcomes of experiments? If not, are resource theorists doing physics?

I will explore answers to these questions. I will introduce TRTs and their role in *one-shot statistical mechanics*, the analysis of work, heat, and entropies on small scales. I will discuss whether TRTs can be tested and whether physicists should care. I will identify eleven opportunities for stepping TRTs closer to experiments. Three opportunities concern what we should realize physically and how, in principle, we can realize it. Six adjustments to TRTs could improve TRTs’ realism. Two opportunities, less critical to realizations, can diversify the platforms with which we might realize TRTs.

The discussion’s broadness evokes a caveat of Arthur Eddington’s. In 1927, Eddington presented Gifford Lectures entitled *The Nature of the Physical World*. Being a physicist, he admitted, “I have much to fear from the expert philosophical critic” [14]. Specializing in TRTs, I have much to fear from the expert experimental critic. This paper is intended to point out, and to initiate responses to, the lack of physical realizations of TRTs. Some concerns are practical; some, philosophical. I expect and hope that the discussion will continue.

## II. TECHNICAL INTRODUCTION

The resource-theory framework is a tool applied to quantum-information problems. Upon introducing the framework, I will focus on its application to statistical mechanics, on TRTs. A combination of TRTs and *one-shot information theory* describes small-scale statistical mechanics. I will introduce select TRT results; more are overviewed in [15].

### A. Resource theories

If you have lived in California during a drought, or fought over the armchair closest to a fireplace, you understand what resources are. *Resources* are quantities that have value. Quantities have value when they are scarce, when limitations restrict the materials we can access and the operations we can perform. *Resource theories* are simple models used to quantify the values ascribed to objects and to tasks by an agent who can perform only certain operations [16]. We approximate the agent as able to perform these operations, called *free operations*, without incurring any cost.

Which quantities have value depends on context. Different resource theories model different contexts, associated with different classes of free operations. Classical agents, for example, have trouble creating entanglement. We can model them as able to perform only local operations and classical communications (LOCC). LOCC define resource theories for entanglement. The resource theory for pure bipartite entanglement is arguably the first, most famous resource theory [17].<sup>1</sup> That theory has been used to quantify the rate at which an agent can convert copies of a pure quantum state  $|\psi\rangle$  into (maximally entangled) Bell pairs. Using a Bell pair and LOCC, one can simulate a quantum channel. Resource states and free operations can simulate operations outside the restricted class.

---

<sup>1</sup>Conventional thermodynamics, developed during the 1800s, is arguably the first, most famous resource theory. But thermodynamics was cast in explicitly resource-theoretic terms only recently.

Resource theorists study transformations between states, such as the quantum states  $\rho$  and  $\sigma$  or the probability distributions  $P$  and  $Q$ . Can the agent transform  $\rho$ , resource theorists ask, into  $\sigma$  via free operations? If not, how much “resourcefulness” does the transformation cost? How many copies of  $\sigma$  can the agent obtain from  $\rho^{\otimes n}$ ? Can  $\rho$  transform into a state  $\tilde{\sigma}$  that resembles  $\sigma$ ? How efficiently can the agent perform information-processing tasks? What is possible, and what is impossible?

So successfully has the resource theory for pure bipartite entanglement answered these questions, quantum information scientists have cast other problems as resource theories. Examples include resource theories for asymmetry [18–20], for stabilizer codes in quantum computation [21], for coherence [22], and for randomness [16]. I will focus on resource theories for thermodynamics.

## B. Thermodynamic resource theories

Given access to a large equilibrated system, an agent ascribes value to out-of-equilibrium systems. Consider a temperature- $T$  heat bath and a hot ( $T' \gg T$ ) gas. By letting the gas discharge heat into the bath, the agent could extract work from equilibration.<sup>2</sup> This work could be stored in a battery, could fuel a car, could power a fan, etc. The resource (the out-of-equilibrium gas) and free operations (equilibration) enable the agent to simulate nonfree operations (to power a car). This thermodynamic story has the skeleton of a resource theory.

The most popular thermodynamic resource theory features an agent given access to a heat bath [1]. The bath has some inverse temperature  $\beta := \frac{1}{k_B T}$ , wherein  $k_B$  denotes Boltzmann’s constant. To specify a state  $R$ , one specifies a density operator  $\rho$  and a Hamiltonian  $H$  defined on the same Hilbert space:  $R = (\rho, H)$ . (Hamiltonians are often assumed to have bounded, discrete spectra.) To avoid confusion between  $R$  and  $\rho$ , I will refer to the former as a *state* and to the latter as a *quantum state*.

The free operations, called *thermal operations*, tend to equilibrate states. Each thermal operation consists of three steps: (i) The agent draws from the bath a Gibbs state  $G$  relative to  $\beta$  and relative to any Hamiltonian  $H_b$ :  $G = (e^{-\beta H_b}/Z, H_b)$ , wherein the partition function  $Z := \text{Tr}(e^{-\beta H_b})$  normalizes the state. (ii) The agent implements any unitary that commutes with the total Hamiltonian:

$$[U, H_{\text{tot}}] = 0, \quad \text{wherein} \quad H_{\text{tot}} := H + H_b = (H \otimes \mathbb{1}) + (\mathbb{1} \otimes H_b). \quad (1)$$

This commutation represents energy conservation, or the First Law of Thermodynamics. (iii) The agent discards any subsystem  $A$  associated with its own Hamiltonian. Each thermal operation has the form

$$(\rho, H) \mapsto \left( \text{Tr}_A \left( U \left[ \rho \otimes \frac{e^{-\beta H_b}}{Z} \right] U^\dagger \right), H + H_b - H_A \right). \quad (2)$$

Each thermal operation decreases or preserves the distance between  $R$  and an equilibrium state (for certain definitions of “distance” [2]). As free operations tend to equilibrate states, and as equilibrium states are free, nonequilibrium states are resources. From nonequilibrium states, agents can extract work. Work is defined in TRTs in terms of batteries, or work-storage systems. A battery can be modeled with a two-level *work bit*  $B_E = (|E\rangle\langle E|, W|W\rangle\langle W|)$  that occupies an energy eigenstate  $|E\rangle \in \{|0\rangle, |W\rangle\}$ . By “How much work can be extracted from  $R$ ?” thermodynamic resource theorists mean (in relaxed notation) “What is the greatest value of  $W$  for which some thermal operation transforms  $R + B_0$  into (*Any state*) +  $B_W$ ?” Answers involve one-shot information theory.

## C. One-shot statistical mechanics

Statistical mechanics involves heat, work, and entropy. Conventional statistical mechanics describes averages over many trials and over many particles. Yet heat, work, and entropy characterize the few-particle scales that experimentalists can increasingly control (e.g., [23–27]), as well as single trials. Small scales have been described by TRTs (e.g., [2, 28–33]). These results fall under the umbrella of *one-shot statistical mechanics*, an application of the *one-shot information theory* that generalizes *Shannon theory* [34].

I will introduce Shannon theory and the *Shannon entropy*  $H_S$  with the thermodynamic protocol of compressing a gas. The quantum counterpart of  $H_S$  is the *von Neumann entropy*  $H_{\text{vN}}$ . Both entropies, I will explain, quantify efficiencies in the large-scale *asymptotic limit* of information theory, which relates to the *thermodynamic limit* of

---

<sup>2</sup>More precisely, the agent could extract the capacity to perform work. I will omit the extra words for brevity.

statistical mechanics. Outside these limits, one-shot entropic quantities quantify efficiencies. I will introduce these quantities, then illustrate with work performance.

*Shannon theory* concerns averages, over many copies of a random variable or over many trials, of the efficiencies with which information-processing tasks can be performed [35]. Many average efficiencies are functions of the Shannon entropy  $H_S$  or the von Neumann entropy  $H_{vN}$ . Examples include the performance of work.

Consider a classical gas in a cylinder capped by a piston. Suppose that the gas, by exchanging heat through the cylinder with a temperature- $T$  bath, has equilibrated. Imagine compressing the gas quasistatically (infinitely slowly) from a volume  $V_i$  to a volume  $V_f$ . One must perform work against the gas particles that bounce off the piston. The work required per trial averages to the change  $\Delta F$  in the gas's *Helmholtz free energy*:

$$\langle W \rangle = \Delta F, \quad \text{wherein} \quad F = \langle E \rangle - TS. \quad (3)$$

$\langle E \rangle$  denotes the average, over infinitely many trials, of the gas's internal energy; and  $S$  denotes the gas's statistical mechanical entropy [36].

The statistical mechanical entropy of a classical state has been cast in terms of the information-theoretic Shannon entropy [36, 37]. The *Shannon entropy* of a discrete probability distribution  $P = \{p_i\}$  is defined as

$$H_S(P) := - \sum_i p_i \ln(p_i). \quad (4)$$

The Shannon entropy quantifies our ignorance about a random variable  $X$  whose possible outcomes  $x_i$  are distributed according to  $P$ . The  $P$  relevant to the gas is the distribution over the microstates that the gas might occupy:  $P = \{e^{-\beta E_i}/Z\}$ , wherein  $E_i$  denotes the energy that the gas would have in Microstate  $i$  and the partition function  $Z$  normalizes the state. The *surprise*  $-\ln(p_i)$  quantifies the information we gain upon discovering that the gas occupies Microstate  $i$ . The surprise quantifies how much the discovery shocks us. Imagine learning which microstate the gas occupies in each of  $n$  trials.  $H_S(P)$  equals the average, in the limit as  $n \rightarrow \infty$ , of the per-trial surprise. The statistical mechanical entropy  $S = k_B H_S(P)$  is proportional to the average of our surprise.  $H_S$  quantifies the average, over  $n \rightarrow \infty$  trials, of the efficiency with which a gas can be compressed quasistatically.

Statistical mechanical averages over  $n \rightarrow \infty$  trials are related to the *thermodynamic limit*, which is related to the *asymptotic limit* of information theory. The thermodynamic limit is reached as a system's volume and particle number diverge ( $V, N \rightarrow \infty$ ), while  $V/N$  remains constant. Imagine performing the following steps  $n$  times: (i) Prepare a system characterized by a particular volume  $V$  and particle number  $N$ . (ii) Measure a statistical mechanical variable, such as the energy  $E$ . Let  $E_\gamma$  denote the outcome of the  $\gamma^{\text{th}}$  trial. Let  $\langle E \rangle$  denote the average, over  $n \rightarrow \infty$  trials, of  $E$ . In the thermodynamic limit, the value assumed by  $E$  in each trial equals the average over trials:  $E_\gamma \rightarrow \langle E \rangle$  [38].

The thermodynamic limit of statistical mechanics relates to an asymptotic limit of information theory. In information theory, the Shannon entropy quantifies averages over  $n \rightarrow \infty$  copies of a probability distribution  $P = \{p_i\}$ . A random variable  $X$  can have a probability  $p_i$  of assuming value the  $x_i$ . Consider compressing  $n$  copies of  $X$ , jointly, into the fewest possible bits. These  $X$ 's are called *i.i.d.*, or *independent and identically distributed*: no variable influences any other, and each variable is distributed according to  $P$ . In the *asymptotic limit* as  $n \rightarrow \infty$ , the number of bits required per copy of  $X$  approaches  $H_S(P)$ . Functions of  $H_S$  quantify averages of the efficiencies with which many classical tasks can be performed in the asymptotic limit [35].

The asymptotic average efficiencies of many quantum tasks depend on the *von Neumann entropy*  $H_{vN}$  [39]. Let  $\rho$  denote a density operator that represents a system's quantum state. The von Neumann entropy

$$H_{vN}(\rho) := -\text{Tr}(\rho \log(\rho)) \quad (5)$$

has an operational interpretation like the Shannon entropy's: Consider preparing  $n$  copies of  $\rho$ . Consider projectively measuring each copy relative to the eigenbasis of  $\rho$ . The von Neumann entropy quantifies the average, in the limit as  $n \rightarrow \infty$ , of our surprise about one measurement's outcome [39].

$H_S$  and  $H_{vN}$  describe asymptotic limits, but infinitely many probability distributions and quantum states are never processed in practice. Nor are infinitely many trials performed. Into how few bits or qubits can you compress finitely many copies of  $P$  or  $\rho$ ? Can you compress into fewer by allowing the compressed message to be decoded inaccurately? How much work must you invest in finitely many gas compressions, e.g., in one "shot"? Such questions have been answered in terms of the *order- $\alpha$  Rényi entropies*  $H_\alpha$  and *Rényi divergences*  $D_\alpha$  [40, 41].

$H_\alpha$ , parameterized by  $\alpha \in [0, \infty)$ , generalizes the Shannon and von Neumann entropies. In the limit as  $\alpha \rightarrow 1$ ,  $H_\alpha(P) \rightarrow H_S(P)$  (if the argument is a probability distribution), and  $H_\alpha(\rho) \rightarrow H_{vN}(\rho)$  (if the argument is a quantum state). Apart from  $H_1$ , two Rényi entropies will dominate our discussion:

$$H_\infty = -\log(p_{\max}) \quad (6)$$

depends on the greatest probability  $p_{\max}$  in a distribution  $P$  or on the greatest eigenvalue  $p_{\max}$  of a quantum state  $\rho$ .  $H_\infty$  relates to the work that a TRT agent must invest to create a state  $(\rho, H)$  [2, 29]. The work extractable from  $(\rho, H)$  relates to

$$H_0 = \log(d) \quad (7)$$

wherein  $d$  denotes the support of  $P$  (the number of nonzero  $p_i$ 's) or the dimension of the support of  $\rho$  (the number of nonzero eigenvalues of  $\rho$ ) [2, 29]. The Rényi divergences  $D_\alpha$  generalize the relative entropy  $D_1$ . They quantify the discrepancy between two probability distributions or two quantum states [39, 41].

Variations on the  $H_\alpha$  and  $D_\alpha$  have been defined and applied (e.g., [34, 42–44]). Examples include the *smooth Rényi entropies* [34]  $H_\alpha^\epsilon$  and the *smooth Rényi divergences*  $D_\alpha^\epsilon$ . I will not define them, to avoid technicalities. But I will discuss them in Sec. IIIB 4 and Sec. IIIB 5, after motivating smoothing with efficiency.

One-shot entropic quantities quantify the efficiencies with which “single shots” of tasks can be performed. Information-processing examples include data-compression rates [34] and channel capacities [45]. Quantum-information tasks include quantum key distribution [34], the distillation of Bell pairs from pure bipartite entangled states  $|\psi\rangle$ , and the formation of  $|\psi\rangle$  from Bell pairs [46]. Thermodynamic efficiencies include the minimum amount  $W_{\text{cost}}$  of work required to create a state  $R$  and the most work  $W_{\text{yield}}$  extractable from  $R$  [2].

The  $W_{\text{cost}}$  and  $W_{\text{yield}}$  of quasiclassical states have been calculated with TRTs. A *quasiclassical* state  $R = (\rho, H)$  has a density operator that commutes with its Hamiltonian:<sup>3</sup>  $[\rho, H] = 0$ . According to [2], the work extractable from one copy of  $R$  is<sup>4</sup>  $W_{\text{yield}}(R) = k_B T D_0(\rho \parallel e^{-\beta H}/Z)$ . Creating one copy of  $R$  costs  $W_{\text{cost}}(R) = k_B T D_\infty(\rho \parallel e^{-\beta H}/Z)$ .

One-shot efficiencies have been generalized to imperfect protocols whose outputs resemble, but do not equal, the desired states [34]. Suppose that a TRT agent wants to create  $R$ . The agent might settle for creating any  $\tilde{R} = (\tilde{\rho}, H)$  whose density operator  $\tilde{\rho}$  is close to  $\rho$ . That is, the agent wants for  $\tilde{\rho}$  to lie within a distance  $\epsilon \in [0, 1]$  of  $\rho$ :  $\mathcal{D}(\tilde{\rho}, \rho) \leq \epsilon$ . (Distance measures  $\mathcal{D}$  are discussed in Sec. IIIB 4.)  $\epsilon$  is called a *smoothing parameter*, *error tolerance*, or *failure probability*. The least amount of work required to create any  $\tilde{R}$  depends on the smooth order- $\infty$  Rényi divergence  $D_\infty^\epsilon$  [43]:  $W_{\text{cost}}^\epsilon(R) = k_B T D_\infty^\epsilon(\rho \parallel e^{-\beta H}/Z)$ . The most work extractable from  $R$  with a similar faulty protocol depends on the smooth order-0 Rényi divergence:  $W_{\text{yield}}^\epsilon(R) = k_B T D_0^\epsilon(\rho \parallel e^{-\beta H}/Z)$  [2].

Let us compare these one-shot work quantities with conventional thermodynamic work. Compressing a gas in the example above costs an amount  $W_{\text{cost}}^{\text{th}} = \Delta F$  of work. Consider expanding the gas from  $V_f$  to  $V_i$ . The gas performs an amount  $W_{\text{yield}}^{\text{th}} = \Delta F$  of work on the piston. In thermodynamics, the work cost equals the extractable work. In one-shot statistical mechanics,  $W_{\text{cost}}(R)$  is proportional to  $D_\infty$ , whereas  $W_{\text{yield}}(R)$  is proportional to  $D_0$ . From the one-shot results, the thermodynamic results have been recovered [2].

### III. OPPORTUNITIES IN PHYSICALLY REALIZING THERMODYNAMIC RESOURCE THEORIES

I have collected eleven opportunities from conversations, papers, and contemplation. Opportunity one concerns the question “Which aspects of TRTs merit realization?” Two and three concern how, in principle, these aspects can be tested. I call for expansions of the TRT framework, intended to enhance TRTs’ realism, in the next six sections. I invite more-adventurous expansions in the final two sections.

#### A. What merits realization? How, in principle, can we realize it?

*Prima facie*, single-particle experiments exemplify the one-shot statistical mechanics developed with TRTs. But many-body systems could facilitate tests. I explain how in section one. Section two concerns our inability to realize the optimal efficiencies described by TRT theorems. Section three concerns which steps one should perform in a lab, and how one should process measurements’ outcomes, to realize TRT results.

##### 1. What would epitomize realizations of the one-shot statistical mechanics developed with TRTs?

Many TRT results fall under the umbrella of *one-shot statistical mechanics*, said to describe small scales. In conventional statistical mechanics, small scales involve few particles and small volumes. Yet much of the mathematics

<sup>3</sup>Many TRT arguments rely on quasiclassicality as a simplifying assumption. After proving properties of quasiclassical systems, thermodynamic resource theorists attempt generalizations to coherent states.

<sup>4</sup> $D_0$  and  $D_\infty$  are called  $D_{\min}$  and  $D_{\max}$  in [2]. I follow the naming convention in [29]: if  $P$  denotes a  $d$ -element probability distribution and  $u$  denotes the uniform distribution  $(\frac{1}{d}, \dots, \frac{1}{d})$ , then  $D_\infty(P||u) = \log(d) - H_\infty(P)$ , and  $D_0(P||u) = \log(d) - H_0(P)$ .

of one-shot statistical mechanics can describe many particles and large volumes. To realize TRT results physically, we must clarify what needs realizing. I will first sketch why one-shot statistical mechanics seems to describe single particles. After clarifying in detail, I will demonstrate how one-shot statistical mechanics can describe many-body systems. I will argue that physically realizing one-shot statistical mechanics involves observing discrepancies between Rényi entropic quantities. This argument implies that entangled many-body states could facilitate physical realizations of several one-shot results.

At first glance, “one-shot” appears to mean “single-particle” in TRT contexts. Conventional statistical mechanics describes collections of about  $10^{24}$  particles. Most of such particles behave in accordance with averages. (Average behavior justifies our modeling of these collections with statistics, rather than with Newtonian mechanics.) The Shannon and von Neumann entropies ( $H_1$ ) quantify averages over large numbers. Other Rényi entropies ( $H_{\alpha \neq 1}$ ) quantify work in one-shot statistical mechanics. Hence one-shot statistical mechanics might appear to describe single particles.

What one-shot statistical mechanics describes follows from what one-shot information theory describes, since the former is an application of the latter. One-shot information theory governs the simultaneous processing of arbitrary numbers  $n$  of copies of a probability distribution  $P$  or of a quantum state  $\rho$ . Hence one-shot statistical mechanics could describe few particles if one copy of  $R = (\rho, H)$  described each particle.

One copy can, but need not, describe one particle. Qubits exemplify the “can” statement. The *qubit*, the quantum analog of the bit, is a unit of quantum information. Qubits form the playground in which many one-shot-statistical-mechanics theorems are illustrated (e.g., [4, 47]). A qubit can incarnate as the state of a two-level quantum system. For example, consider an electron in a magnetic field  $\mathbf{B} = B\hat{z}$ . The Zeeman Hamiltonian  $H_Z \propto \sigma_z$  governs the electron’s spin, wherein  $\sigma_z$  denotes the Pauli  $z$ -operator.  $H_Z$  defines two energy levels, and a density operator  $\rho$  represents the electron’s spin state. If  $R = (\rho, H_Z)$ , the one-shot quantities  $W_{\text{cost}}(R)$  and  $W_{\text{yield}}(R)$  represent the work cost of creating, and the work extractable from, one electron. Hence one copy of a state  $R$  can describe one particle; and, as argued above, one-shot statistical mechanics can describe single particles.

Yet one copy of  $R$  can represent the state of a system of many particles. For example, let  $R = (\rho, H_{XX})$  denote the state of a chain of spin- $\frac{1}{2}$  particles. Such a chain is governed by the Hamiltonian  $H_{XX} = -\frac{J_{\text{ex}}}{2} \sum_j (S_j^+ S_{j+1}^- + S_{j+1}^+ S_j^-)$ , wherein  $J_{\text{ex}}$  denotes the uniform coupling’s strength; the index  $j$  can run over the arbitrarily many lattice sites; and  $S_j^+$  and  $S_j^-$  denote the  $j^{\text{th}}$  site’s raising and lowering operators [48]. This many-body state  $R$  can obey TRT theorems derived from one-shot information theory. Hence one-shot statistical mechanics can describe many-body systems.

My point is not that one-shot functions can be evaluated on  $n > 1$  i.i.d. copies of a state, denoted by  $R^n = (\rho^{\otimes n}, \oplus_{i=1}^n H)$ . One-shot functions, such as  $H_\alpha$  or  $W_{\text{cost}}$ , are evaluated on  $R^n$  in arguments about the asymptotic limit. Asymptotic limits of one-shot functions lead to expressions reminiscent of conventional statistical mechanics [e.g., Eqs. (3)], which describes many particles [2, 29, 32]. Therefore,  $R^n$  is known to be able to represent the state of a many-particle system. But my point is that  $R$ , not only  $R^n$ , can represent the state of a many-particle system.  $W_{\text{cost}}(R)$  can equal the work cost of creating “one shot” of a many-particle state [30, 31].

Therefore, few-particle systems would not necessarily epitomize physical realizations of one-shot statistical mechanics. What would? Quantum states  $\rho$  whose one-shot entropic quantities ( $D_{\alpha \neq 1}$ ,  $H_0^\epsilon$ ,  $H_\infty^\epsilon$ , etc.) were nonzero and differed greatly from asymptotic entropic quantities ( $D_1$  and  $H_1$ ) and from each other [49]. I will illustrate with the “second laws” for coherence and with quasiclassical work.

A coherence property of a quantum state  $R$ , called the *free coherence*, has been quantified with the Rényi divergences  $D_\alpha$  [32]. A thermal operation can map  $R$  to  $S$  only if the  $D_\alpha$  decrease monotonically for all  $\alpha \geq 0$ . In the asymptotic limit, the average free coherence per particle vanishes. Schematically,  $D_\alpha/n \rightarrow 0$  [32]. Observing coherence restrictions on thermodynamic evolutions—observing a TRT prediction—involves observing nontrivial Rényi divergences.

Second, the approximate work cost  $W_{\text{cost}}^\epsilon(R)$  of a quasiclassical state  $R = (\rho, H)$  depends on  $D_\infty^\epsilon(\rho || e^{-\beta H}/Z)$ , and the approximate work yield  $W_{\text{yield}}^\epsilon(R)$  depends on  $D_0^\epsilon(\rho || e^{-\beta H}/Z)$ . In conventional statistical mechanics, the work cost and the work yield depend on the same entropic quantity. Physically realizing one-shot statistical mechanics involves observing discrepancies between Rényi divergences.

We might increase our chances of observing such a discrepancy by processing a quantum state  $\rho$  whose Rényi entropies differ greatly. Many-body entangled systems can occupy such states [49]. Example states include the generalization

$$|\psi_n^W\rangle = \frac{1}{\sqrt{n}}(|1 \underbrace{0 \dots 0}_{n-1}\rangle + |010 \dots 0\rangle + \dots + |0 \dots 01\rangle) \quad (8)$$

of the W state [49]. Whereas  $H_\infty^\epsilon(|\psi_n^W\rangle\langle\psi_n^W|) = 0$ ,  $H_0^\epsilon(|\psi_n^W\rangle\langle\psi_n^W|) = n$ . If  $n$  is large, these entropies differ greatly. Not only can one-shot statistical mechanics describe many-body systems, but entangled many-body states might facilitate tests of one-shot theory.

Conversely, one-shot theory might offer insights into entangled many-body states. Such states have been called “exotic,” and engineering them poses challenges. Suppose that thermal operations modeled the operations performable

in a condensed-matter or optics lab. One-shot statistical mechanics would imply which states an experimentalist could transform into which, how much energy must be drawn from a battery to create a state, which coherences could transform into which, etc.

Not all one-shot results could be realized with large systems. A few results depend on the dimensionality of the Hilbert space on which a quantum state is defined. An example appears in [33, App. G.4].<sup>5</sup>

## 2. How can we test predictions of maximal efficiencies?

Many TRT theorems concern the maximal efficiency with which any thermal operation can implement some transformation  $R \mapsto S$ . “Maximal efficiency” can mean “the most work extractable from the transformation,” for example. Testing maximal efficiencies poses two challenges: (i) No matter how efficiently one implements a transformation, one cannot know whether a more efficient protocol exists. (ii) Optimal thermodynamic protocols tend to last for infinitely long times and to involve infinitely large baths. These challenges plague not only TRTs, but also conventional thermodynamics. We can approach the challenges in at least three ways: (A) Deprioritize experimental tests. (B) Calculate corrections to the predictions. (C) Check whether decreasing speeds and increasing baths’ sizes inches realized efficiencies toward predictions.

Proving that one has implemented a transformation maximally efficiently would amount to proving a negative. Experiments cannot prove such negatives. Imagine an agent who wishes to create some quasiclassical state  $R$  from a battery and thermal operations. Suppose that creating any state  $\tilde{R}$  that lies within a distance  $\epsilon$  of  $R$  would satisfy the agent. The least amount  $W_{\text{cost}}^\epsilon(R)$  of work required to create any such  $\tilde{R}$  was calculated in [2]. In what fashion could an experimentalist test this prediction, if able to perform arbitrary thermal operations?

The form of a thermal operation that generates a  $\tilde{R}$  is implied in [2].<sup>6</sup> One could articulate the operation’s form explicitly and could perform the operation in each of many trials. If  $\epsilon$ -closeness is defined in terms of the trace distance, one could measure each trial’s output, then calculate the distance between the created state and  $R$ . One could measure the work  $W$  invested in each trial and could check whether  $W = W_{\text{cost}}^\epsilon(R)$ . Separately, one could create  $\tilde{R}$ ’s from another thermal operation, could measure the work  $W$  invested in each trial, and could check whether each  $W > W_{\text{cost}}^\epsilon(R)$ . Finally, one could design a thermal operation  $\mathcal{E}$  that outputs a  $\tilde{R}$  if sufficient work is invested, could invest  $W < W_{\text{cost}}^\epsilon(R)$  in a realization of  $\mathcal{E}$ , and could check that the resulting state differs from  $R$  by more than  $\epsilon$ . Yet one would not know whether, by investing  $W < W_{\text{cost}}^\epsilon(R)$  in another protocol, one could create a  $\tilde{R}$ .

Testing optima poses problems also because optimal thermodynamic protocols tend to proceed quasistatically. Quasistatic, or infinitely slow, processes keep a system in equilibrium. Quick processes tend to eject a system from equilibrium, dissipating extra heat [4]. This heat, by the First Law, comes from work drained from the battery [37]. Hence TRT predictions about optimal efficiencies are likely predictions of the efficiencies of quasistatic protocols. Quasistatic protocols last for infinitely long times. Long though graduate students labor in labs, their experiments do not last forever.

Baths’ sizes impede experimental tests as time does. In [2, Suppl. Note 1], “the energy of the heat bath (and other relevant quantities such as [the] size of degeneracies) [tends] to infinity.” Real baths have only finite energies and degeneracies. Granted, baths are assumed to be infinitely large (and unrealistically Markovian) in statistical mechanics outside of TRTs. As such predictions must be reconciled with real baths’ finiteness, so must TRT predictions.

The reconciliation can follow one of at least three paths. First, resource theorists can shrug. Shruggers would have responded to the question at my seminar with “No, my arch-nemesis (new friend) might as well not have heard of my results, for all they’ll impact his lab.” The  $W_{\text{cost}}^\epsilon$  prediction appears in the paper “Fundamental limitations for quantum and nanoscale thermodynamics.” Fundamental limitations rarely impact experiments more than the imperfectness of experimentalists’ control does. For example, quantum noise fundamentally impedes optical amplifiers [50]. Johnson noise, caused by thermal fluctuations, impedes amplifiers in practice, not fundamentally. That is, cooling an amplifier can eliminate Johnson noise. Yet Johnson noise often outshouts quantum noise. As we cannot always observe fundamental, quantum noise, we should not necessarily expect to observe the fundamental limitations predicted with TRTs.

Nor, one might continue, need we try to observe fundamental limitations. Fundamental limitations bound the efficiencies of physical processes. Ideal bounds on achievable quantities have been considered physics. Thermodynamics

<sup>5</sup>The authors discuss *catalysis*, the use of an ancilla to facilitate a transformation. Let  $R = (\rho, H_R)$  denote a state that cannot transform into  $S = (\sigma, H_S)$  by thermal operations:  $R \not\mapsto S$ . Some *catalyst*  $C = (\xi, H_C)$  might satisfy  $(\rho \otimes \xi, H_R + H_C) \mapsto (\sigma \otimes \xi, H_S + H_C)$ . Catalysts act like engines used to extract work from a pair of heat baths. Engines degrade, so a realistic transformation might yield  $\sigma \otimes \tilde{\xi}$ , wherein  $\tilde{\xi}$  resembles  $\xi$ . For certain definitions of “resembles,” the agent can extract arbitrary amounts of work by negligibly degrading  $C$ . Brandão *et al.* quantify this extraction in terms of the dimension  $\dim(\mathcal{H}_C)$  of the Hilbert space  $\mathcal{H}_C$  on which  $\xi$  is defined. The more particles the catalyst contains, the greater the  $\dim(\mathcal{H}_C)$ . Such one-shot results depend on the number of particles in the system represented by “one copy” of  $C$ .

<sup>6</sup>The functional form of  $W_{\text{cost}}^\epsilon(R)$  is derived in [2, Suppl. Note 4]. The proof relies on Theorem 2. The proof of Theorem 2 specifies how a TRT agent can perform an arbitrary free unitary. Hence the thermal operation that generates a  $\tilde{R}$  is described indirectly.

counts as physics, though some thermodynamic transformations are quasistatic. TRTs need no experimental tests; ignore the rest of this paper.

Yet thermodynamics has withstood tests [51]. “Quantum-limited” amplifiers uninhibited by Johnson noise have been constructed. Fundamental limitations have impacted experiments, so fundamental limitations derived from TRTs merit testing. Furthermore, testable predictions distinguish physics from philosophy and mathematics. If thermodynamic resource theorists are doing physics, observing TRT predictions would be fitting. Finally, resource theorists motivate TRTs partially with experiments. Experimentalists, according to one argument, can control single molecules and can measure minuscule amounts of energy. Conventional statistical mechanics models such experiments poorly. Understanding these experiments requires tools such as TRTs. Do such arguments not merit testing? If experimentalists observe the extremes predicted with TRTs, then the justifications for, and the timeliness of, TRT research will grow.

Tests of TRT optima can be facilitated by the calculation of corrections and by experimental approaches to ideal conditions. To compute corrections, one could follow Reeb and Wolf [52]. They derive corrections, attributable to baths’ finiteness, to Landauer’s Principle. [Landauer proposed that erasing one bit of information quasistatically costs an amount  $W_L = k_B T \ln(2)$  of work.] The authors’ use of quantum-information tools could suit the resource-theory framework. One-shot tools are applied to finite-size baths in [53], and the resource-theory framework is applied in [54].

Instead of correcting idealizations theoretically, one might approach idealizations experimentally. One can perform successive trials increasingly slowly, and one can use larger and larger baths. As speeds drop and baths grow, do efficiencies approach the theoretical bound? Koski *et al.* posed this question about speed when testing Landauer’s Principle. Figure 2 in [55] illustrates how slowing a protocol drops the erasure’s cost toward  $W_L$ . Approaches to the quasistatic limit and to the infinitely-large-bath limit might be used to test TRTs.

### 3. Which operations should be performed to test TRT results?

To test a theorem about a transformation, one should know how to implement the transformation and what to measure. Consider a prediction of the amount of work required to transform a state  $R$  into a state  $S$ . Some thermal operation  $\mathcal{E}$  satisfies  $\mathcal{E}(R) = S$ . Checking the prediction requires knowledge of the operation’s form. Which Gibbs state must one create, which unitary must one implement, and which subsystem must one discard? What must be measured in how many trials, and how should the measurement outcomes be processed?

Proofs of TRT theorems detail thermal operations to different extents. Less-detailed proofs include that of Theorem 5 in [30]. The authors specify how a channel transforms a state by specifying the form of the channel’s output. Which Gibbs state, unitary, and tracing-out the channel involves are not specified. Farther up the explicitness spectrum lies the proof that, at most, an amount  $W_{\text{yield}}^\epsilon(R)$  of work can be extracted with accuracy  $1 - \epsilon$  from a quasiclassical state  $R$  [2, Suppl. Note 4]. The proof specifies that “we have to map strings of [weight  $1 - \epsilon$ ] to a subspace of our energy block.” The agent’s actions are described, albeit abstractly. The same proof specifies another operation in greater detail, in terms of unitaries: “For each fixed [energy]  $E_S$  [of the system of interest,] we apply a random unitary to the heat bath, and identity to [the system of interest]” [2, Suppl. Note 2].

One could describe an operation’s form more explicitly. For instance, one could specify how strong a field of which type should be imposed on which region of a system for how long. Such details might depend on one’s platform—on whether one is manipulating a spin chain, ion traps, quantum dots, etc. Such explicitness would detract from the operation’s generality. Generality empowers TRTs. Specifying a Gibbs state, a unitary, and a tracing-out would balance generality with the facilitation of physical realizations.

In particular need of unpacking are the clock, catalyst, coupling, and time-evolution formalisms. Resource theorists developed each formalism, in a series of papers, to model some physical phenomenon. Later papers, borrowing these formalisms, reference the earlier papers implicitly or briefly. The scantiness of these references expedites theoretical progress but can mislead those hoping to test the later results. I will overview, and provide references about, each formalism.

The clock formalism is detailed in [1, 2, 56]. Many TRT theorems concern transformations  $(\rho, H) \mapsto (\sigma, H)$  that preserve the Hamiltonian  $H$ . Theorems are restricted to constant Hamiltonians “without loss of generality.” Yet Hamiltonians change in real-life protocols, as when fields  $\mathbf{B}(t)$  are quenched. Resource theorists reconcile the changing of real-life Hamiltonians with the frozenness of TRT Hamiltonians by supposing that a clock couples to the system of interest. When the clock occupies the state  $|i\rangle$ , the Hamiltonian  $H(t_i)$  governs the system of interest. The composite system’s Hamiltonian remains constant, so TRT theorems describe the composite. TRT theorems restricted to constant  $H$ ’s owe their generality to the clock formalism.

Like the clock formalism, catalysis requires detailing. Laboratory equipment such as clocks facilitates experiments without changing much. These items serve as *catalysts*, according to the resource-theory framework. If TRT predictions are tested in some lab, the catalysts in the lab must be identified. Predictions should be calculated from catalytic



thermal operations [28, 33]. If predictions are calculated from thermal operations, the neglect of the catalysts must be justified.

*Prima facie*, couplings do not manifest in TRTs. Free unitaries couple subsystems  $R = (\rho, H)$  and  $G = (e^{-\beta H_b}/Z, H_b)$ . No interaction term  $H_{\text{int}}$  appears in  $H_{\text{tot}} = H + H_b$ . But interaction Hamiltonians couple subsystems in condensed matter and quantum optics. Since condensed matter and quantum optics might provide testbeds for TRTs, the coupling formalisms must be reconciled. An equivalence between the formalisms is discussed in [1, Sec. VIII].

*Prima facie*, TRT systems do not evolve in time. Many authors define states  $R = (\rho, H)$  but mention no time evolution of  $\rho$  by  $U(t) = \exp(-\frac{i}{\hbar}Ht)$ . These authors focus on quasiclassical states, whose density operators commute with their Hamiltonians:  $[\rho, H] = 0$ . This  $H$  generates  $U(t)$ . Hence if  $[\rho, U(t)] = 0$ , the states remain constant in the absence of interactions, and time evolution can be ignored. Quantum states  $\rho$  that have coherences relative to the  $H$  eigenbasis evolve nontrivially under their Hamiltonians. This evolution commutes with thermal operations [32, 47]. Though they might appear not to, time evolution and couplings familiar from condensed matter and optics manifest in TRTs.

In addition to the specifying the steps in thermal operations, resource theorists could specify what needs measuring, how precisely, in how many trials, and how to combine the measurements' outcomes, to test TRT predictions. For example, many TRT theorems involve  $\epsilon$ -approximation.  $\epsilon$ -approximation is often defined in terms of the trace distance  $\mathcal{D}_{\text{tr}}$  between quantum states.  $\mathcal{D}_{\text{tr}}$  has an operational interpretation. How to use that interpretation, to check the distance from an experimentally created state to the desired state, is discussed in Sec. III B 4. Instead of invoking the trace distance's operational interpretation, one could perform quantum state tomography. The preciseness with which measurements can be performed must be weighed against the error tolerance  $\epsilon$  in the theorems being tested. As another example, some TRT predictions cannot be tested, as discussed in Sec. III A 2. We must distinguish which predictions are testable; then specify which measurements, implemented how precisely, and combined in which ways, would support or falsify those predictions.

## B. Enhancing TRTs' realism

Six adjustments could improve how faithfully TRTs model reality. First, we could narrow the gap between the operations easily performable by TRT agents and the operations easily performable by thermodynamic experimentalists. Second, we could incorporate violations of energy conservation into TRTs: TRTs model closed systems, whereas most real systems are open. Section III B 3 concerns measurements; and Sec. III B 4 concerns thermal embezzlement, a seeming violation of the First Law of Thermodynamics in TRTs. Faulty operations feature in Sec. III B 5. Finally, modeling continuous spectra would improve TRT models of classical systems and quantum environments.

### 1. Should free operations reflect experimentalists' capabilities better? If so, how?

Should resource theorists care about how easily experimentalists can implement thermal operations? Laboratory techniques advance. Even if thermal operations challenge experimentalists now, they might not in three years. Theory, some claim, waits for no experiment. Quantum error correction (QEC) illustrates this perspective. Proposals for correcting a quantum computer's slip-ups involve elaborate architectures, many measurements, and precise control [57]. QEC theory seemed unrealizable when it first flowered. Yet the foundations of a surface code have been implemented [58], and multiple labs aim to surpass these foundations. Perhaps theory should motivate experimentalists to expedite difficult operations.

On the other hand, resource theories are constructed partially to reflect limitations on labs. The agents in resource theories reflect the operationalism inspired partially by experiments. Furthermore, TRTs were constructed to shed light on thermodynamics. Thermodynamics evolved largely to improve steam engines: thermodynamics evolved from practicalities. By ignoring practical limitations, thermodynamic resource theorists forsake one of their goals.<sup>7</sup> Finally, increasing experimental control over nanoscale systems has motivated thermodynamic resource theorists. Experimentalists, an argument goes, are probing limits of physics. They are approaching regimes near the fundamental limitations described by TRTs. This argument merits testing. Physical realizations of TRTs would strengthen the justifications for developing TRTs.

---

<sup>7</sup>Other goals include the mathematical isolation, quantification, and characterization of single physical quantities. Want to learn how entanglement empowers you to create more states and to perform more operations than accessible with only separable states? Use a resource theory for entanglement. Want to learn how accessing information empowers you? Use the resource theory for information [29, 59].

Thermodynamic experimentalists disagree with resource theorists in three ways about which operations are easy: experimentalists cannot necessarily easily implement unitaries that commute with  $H_{\text{tot}}$ . Nor can experimentalists create arbitrary Hamiltonians. Finally, experimentalists do not necessarily value work as TRT agents do. Redefining thermal operations could remedy these discrepancies.

TRTs agents can perform any unitary  $U$  that commutes with the total Hamiltonian  $H_{\text{tot}} = H + H_b$  of a system-and-bath composite [Eq. (1)]. Thermodynamic experimentalists cannot necessarily. Consider, for example, condensed matter or quantum optics. Controlling long-range interactions and generating many-body interactions can be difficult. Two-body interactions are combined into many-body interactions. Hence many-body interactions are of high order in two-body coupling constants. Particles resist dancing to the tune of such weak interactions. Coaxing the particles into dancing challenges experimentalists but not TRT agents.

In addition to implementing energy-conserving unitaries, a TRT agent can create equilibrium states ( $e^{-\beta H_b}/Z, H_b$ ) relative to the heat bath's inverse temperature  $\beta$  and relative to any Hamiltonian  $H_b$  (though  $H_b$  is often assumed to have a bounded, discrete spectrum). Experimentalists cannot create systems governed by arbitrary  $H_b$ 's. Doing so would amount to fabricating arbitrary physical systems. Though many-body and metamaterials experimentalists engineer exotic states, they cannot engineer everything.

Most relevantly, experimentalists cannot construct infinitely large baths. Let  $\dim(\mathcal{H}_b)$  denote the dimension of the Hilbert space  $\mathcal{H}_b$  on which the bath's quantum state is defined. According to the often-used prescription in [2], the output of a thermal channel approaches the desired state  $R = (\rho, H)$  in the limit as  $\dim(\mathcal{H}_b) \rightarrow \infty$ . Real baths' states are not defined on infinitely large Hilbert spaces. Baths' finiteness merit incorporation into TRTs.

Just as experimentalists cannot construct arbitrary bath Hamiltonians  $H_b$ , experimentalists cannot construct arbitrary ancillary Hamiltonians. Work, for example, is often defined in TRTs with a two-level work bit [2]. Suppose that an agent wishes to transform  $R$  into  $S$  by investing an amount  $W$  of work. The agent borrows a work bit whose gap equals  $W$ :  $B_W = (|W\rangle\langle W|, W|W\rangle\langle W|)$ . An experimentalist should prepare  $B_W$  to mimic the agent. But an experimentalist cannot necessarily tune a two-level system's gap to  $W$ . Resource-theory predictions might require recalculating in terms of experimentally realizable ancillas.

Whereas TRT agents can easily implement unitaries and create states that experimentalists cannot, some thermodynamic experimentalists can easily spend work that TRT agents cannot. For example, consider a TRT agent who increases the strength  $B(t)$  of a magnetic field  $\mathbf{B}(t)$  imposed on a spin system governed by a Hamiltonian  $H(t)$ . The Hamiltonian's evolution is modeled with the clock formalism in [1, 2]. From this formalism, one can calculate the minimum amount  $W_{\text{cost}}$  of work required to increase  $B(t)$ .

The experimentalist modeled by the agent expends work to increase  $B(t)$ . One can strengthen a magnetic field by strengthening the current that flows through a wire. One strengthens a current by heightening the voltage drop between the wire's ends, which involves strengthening the electric field at one end of the wire, which involves bringing charges to that end. The charges present repel the new charges, and overcoming the repulsion requires an amount  $W'_{\text{cost}}$  of work. Yet thermodynamic experimentalists do not necessarily cringe at the work cost of strengthening a field. They take less pride in charging a battery than in engineering many-body and long-range interactions, in turning a field on and off quickly, and in sculpting a field's spatial profile  $\mathbf{B}(\mathbf{r})$ . The work cast by resource theorists as valuable has less value, in some thermodynamic labs, than other resources.

Regardless of whether they value work, experimentalists might be able to measure  $W'_{\text{cost}}$  [27]. This cost can be compared with the predicted value  $W_{\text{cost}}$ . TRT results might be tested even if the theory misrepresents some priorities of some thermodynamic experimentalists.<sup>8</sup>

Resource theorists might incorporate experimentalists' priorities and challenges into TRTs. The set of free unitaries, and the set of Hamiltonians  $H_b$  relative to which agents can create Gibbs states, might be restricted [62]. One might calculate the cost of implementing a many-body interaction. Corrections might be introduced into existing calculations. Additionally, alternative battery models might replace work bits.

Precedents for altering the definition of free operations exists: thermal operations are expanded to *catalytic thermal operations* in [28, 33]. Experimentalists use engines, clocks, and other equipment that facilitates transformations without altering (or while altering negligibly). These catalysts, Brandão *et al.* argue, merit incorporation into free operations. Batteries are similarly incorporated into free operations in [63]. Inspired by [63], Åberg redefined free operations to expose how coherences catalyze transformations [60].

Limitations on experimentalists' control have begun infiltrating TRTs. Brandão *et al.* showed that each of many thermal operations  $\mathcal{E}_1, \mathcal{E}_2, \dots$  can implement  $R \mapsto S$  [1]. If an agent can implement any  $\mathcal{E}_i$ , the agent can perform the transformation. Performing a particular  $\mathcal{E}_i$ —controlling particular aspects of the operations—is unnecessary [1, Suppl. Note 7]. Wilming *et al.* compared TRT agents with agents who can only thermalize states entirely (can only

<sup>8</sup>Such experimentalists value coherence similarly to TRT agents: in experiments, coherent entangled states offer access to never-before-probed physics. In quantum computers, entanglement speeds up calculations. TRT agents value coherence because they can catalyze transformations with coherent states [60]. Agents can also “unlock” work from coherence [61].

replace states with Gibbs states) [64]. Real experimentalists lie between these extremes: they can perform some, but not all, thermal operations apart from complete thermalization. Following these authors, resource theorists might improve the fidelity with which free operations reflect reality.

## 2. Can nonconservation of energy model systems' openness and model the impreciseness with which unitaries are implemented?

TRTs describe closed systems, whose energies are conserved. Every free unitary  $U$  commutes with the total Hamiltonian:

$$[U, H_{\text{tot}}] = 0. \quad (9)$$

Experimental systems are open and have not-necessarily-conserved energies. We might reconcile the closedness of TRT systems with the openness of experimental systems by modifying the constraint (9). A modification could model the impreciseness with which unitaries can be implemented [1].

All subsystems of our universe interact with other subsystems. A quantum system  $\mathcal{S}$  in a laboratory might couple to experimental apparatuses, air molecules, etc. Batteries  $\mathcal{B}$ , clocks  $\mathcal{C}$ , and catalysts  $\mathcal{C}$  have been studied in TRTs. The energy of  $\mathcal{S} + \mathcal{B} + \mathcal{C} + \mathcal{C}$  can more justifiably be approximated as conserved than the energy of  $\mathcal{S}$  can. Yet the energy of  $\mathcal{S} + \mathcal{B} + \mathcal{C} + \mathcal{C}$  can change. If light shines on the clock, photons interact with  $\mathcal{C}$ . We cannot include in our calculations all the degrees of freedom in a closed, conserved-energy system. To compromise, we might incorporate into TRTs the nonconservation of the energies of  $\mathcal{S} + \mathcal{B} + \mathcal{C} + \mathcal{C}$ .

Brandão *et al.* introduced nonconservation of energy into TRTs [1, App. VIII]. The authors suppose that work is extracted from  $\mathcal{S}$  during  $n$  cycles. The noncommutation of  $U$  with  $H_{\text{tot}}$ , they show, corrupts the clock's state. This corruption disturbs the work extraction negligibly in the limit as  $n \rightarrow \infty$ . Even if unable to implement  $U$  precisely, the authors conclude, an agent can extract work effectively. The authors' analysis merits detailing and invites extensions to noncyclic processes.

In addition to modeling lack of control, a relaxation of energy conservation would strengthen the relationship between TRTs and conventional statistical mechanics. As Åberg writes in [60], "it ultimately may be desirable to develop a generalization which allows for small perturbations of the perfect energy conservation, i.e., allowing evolution within an energy shell, as it often is done in statistical mechanics."

## 3. How should TRTs model measurements?

The most general process modeled by quantum information theory consists of a preparation procedure, an evolution, and a measurement [39]. A general measurement is represented by a *positive operator-valued measure* (POVM). A POVM is a set  $\mathcal{M} = \{M_i\}$  of measurement operators  $M_i$ . The probability that some measurement of a quantum state  $\rho$  yields outcome  $i$  equals  $\text{Tr}(\rho M_i)$ .

Measurements are (postselection is) probabilistic, whereas thermal operations are deterministic. The probability that some thermal operation  $\mathcal{E}$  transforms some state  $R = (\rho, H_R)$  into some state  $S = (\sigma, H_S)$  equals one or zero. Not even if supplemented with a battery can thermal operations implement probabilistic transformations. The absence, from TRTs, of the measurement vertebra in the spine of a realistic protocol impedes the modeling of real physical processes.

The absence also impedes the testing of TRTs. We extract data from physical systems by measuring them. We measure quantum systems through probabilistic transformations. If we attribute to a classical statistical mechanical system a distribution over possible microstates, a measurement of the system's microstate is probabilistic. If TRTs do not model measurements, can we test TRTs?

Alhambra *et al.* propose a compromise that I will recast [65]. Suppose that an agent wishes to measure a state  $\rho' = p\sigma + (1-p)X$  of a system  $\mathcal{S}$ , assuming that  $\sigma$  and  $X$  denote density operators. Suppose that the agent has borrowed a memory  $\mathcal{M}$ , initialized to a nonfree pure state  $(|0\rangle\langle 0|, \mathbb{1}_d)$ , from some bank.  $\mathcal{M}$  is governed by a totally degenerate Hamiltonian  $\mathbb{1}_d$  defined on a  $d$ -dimensional Hilbert space. A free unitary  $U_{\text{record}}$  can couple  $\mathcal{S}$  to  $\mathcal{M}$ . Measuring  $\mathcal{M}$  would collapse the state of  $\mathcal{S}$  onto  $\sigma$  or onto  $X$ , implementing a quantum POVM.

To incorporate the memory measurement into TRTs, we might add paid-for measurements of quasiclassical states into the set of allowed operations. The agent should be able to measure a quasiclassical memory, then reset the

memory to some fiducial state by investing work.<sup>9</sup> Schematically,

$$(\rho_M, \mathbb{1}_d) + (|W\rangle\langle W|, W|W\rangle\langle W|) \mapsto (|0\rangle\langle 0|, \mathbb{1}_d) + (|0\rangle\langle 0|, W|W\rangle\langle W|). \quad (10)$$

The agent originally ascribes to the memory a distribution over possible pure states. Observing which microstate  $\mathcal{M}$  occupies, the agent gains information transformable into work [66]. Gaining work deterministically from free operations contradicts the spirit of the Second Law of Thermodynamics. But the agent forfeits this work to erase  $\mathcal{M}$ . The agent must return  $\mathcal{M}$ , restored to its original state, to the bank.<sup>10</sup>

Questions about this measurement formalism remain. First, suppose that  $U_{\text{record}}$  correlates  $\mathcal{S}$  with  $\mathcal{M}$  perfectly. The memory's reduced state is maximally mixed, having the spectrum  $P_{\mathcal{M}} = (\frac{1}{d} \dots \frac{1}{d})$ . Landauer's Principle suggests that erasing  $\mathcal{M}$  costs at least an amount  $W = k_B T \log(d)$  of work [67]. Instead, suppose that  $U_{\text{record}}$  correlates  $\mathcal{S}$  with  $\mathcal{M}$  imperfectly.  $P_{\mathcal{M}}$  might not be maximally mixed; erasing  $P_{\mathcal{M}}$  could cost an amount  $W(P_{\mathcal{M}}) < k_B T \log(d)$  of work [29].

Second, measuring and erasing  $n$  copies of  $\mathcal{M}$  individually costs an amount  $n W(P_{\mathcal{M}})$  of work. But the agent might prefer paying a lump sum to paying for each copy individually. In the limit as  $n \rightarrow \infty$ ,  $(P_{\mathcal{M}})^{\otimes n}$  can be compressed into  $n H_S(P_{\mathcal{M}})$  bits. Delaying payment might save the agent work [68]. Delayed payments, like imperfect  $U_{\text{record}}$ 's, as well as imperfect measurements and nondegenerate memories, merit consideration.

A precedent exists for incorporating paid-for measurements into allowed operations: thermal operations have been expanded to catalytic thermal operations [28, 33]. Just as catalysts are not free, measurements are not: an agent could pay to restore a degraded catalyst to its initial state and can pay to erase a memory. As catalysis has been incorporated into allowed operations, so might measurements. The term  *$\epsilon$ -deterministic thermal operations* has already appeared [5].  $\epsilon$ -determinism surfaces in a related framework, if not in the resource-theory framework, in [3]. Navascués and García-Pintos treat measurement similarly to Alhambra *et al.* when studying the “resourcefulness” of nonthermal operations [69].

On the other hand, physically realizing TRTs might not necessitate the incorporation of measurements into TRTs. An experimenter could test a work-extraction theorem, derived from the TRT framework, as follows: First, the experimenter extracts work by implementing a thermal operation. Then, the experimenter performs a measurement absent from the TRT framework. The experimenter imitates two agents: One, a TRT agent, extracts work by a thermal operation. This first agent passes the resultant system to someone who “lives outside the resource theory,” whom energy constraints do not restrict. The second agent measures the system. Hence one might test TRTs that do not model measurements. But externalizing measurements renders TRTs incomplete models of simple thermodynamic processes. To model physical reality, we must refine the TRT representation of measurements.

#### 4. Can a measure of $\epsilon$ -closeness lead to testable predictions and away from embezzlement?

Many TRT results concern  $\epsilon$ -approximations (e.g., [2, 28–31, 33, 65, 69]): Suppose that an agent wishes to create a *target state*  $R = (\rho, H)$ . The agent might settle for some  $\tilde{R} = (\tilde{\rho}, H)$  whose density operator  $\tilde{\rho}$  lies within a distance  $\epsilon \in [0, 1]$  of  $\rho$ :  $\mathcal{D}(\rho, \tilde{\rho}) \leq \epsilon$ . This  $\epsilon$  has been called the *error tolerance*, *failure probability*, and *smoothing parameter*. The distance measure  $\mathcal{D}$  is often chosen to be the trace distance  $\mathcal{D}_{\text{tr}}$ .  $\mathcal{D}_{\text{tr}}$  has an operational interpretation that could facilitate experimental tests. But  $\mathcal{D}_{\text{tr}}$  introduces *embezzlement* into TRTs. Thermal embezzlement is the extraction of arbitrary amounts of work from the negligible degradation of a catalyst. A *catalyst* consists of equipment, such as an engine, that facilitates a transformation while remaining unchanged or almost unchanged. Negligible degradation—the cost of embezzlement—is difficult to detect. Extracting work at a difficult-to-detect cost contradicts the spirit of the First Law of Thermodynamics. Resource theorists have called for eliminating embezzlement from TRTs by redefining thermal operations or by redefining  $\epsilon$ -approximation [33, 49, 70]. Three redefinitions have been proposed. On the other hand, embezzlement might merit physical realization.

I will begin with background about the trace distance, defined as follows [39]. Let  $\rho$  and  $\tilde{\rho}$  denote density operators defined on the same Hilbert space. The *trace distance* between them is

$$\mathcal{D}_{\text{tr}}(\rho, \tilde{\rho}) := \frac{1}{2} \text{Tr}|\rho - \tilde{\rho}|, \quad (11)$$

<sup>9</sup>One might worry that  $\mathcal{M}$  might not occupy a quasiclassical state after coupling to  $\mathcal{S}$ . But  $\mathcal{M}$  has a totally degenerate Hamiltonian  $\mathbb{1}_d$ . If  $\rho_M$  has coherences relative to the energy eigenbasis, free unitaries can eliminate them [29].

<sup>10</sup>One might object that the measurement could project the memory's quantum state onto a pure state. Transforming any pure state into  $|0\rangle$  costs no work: The transforming unitary commutes with the Hamiltonian  $\mathbb{1}_d$  and so is free [29]. But the agent could be fined the work that one would need to erase  $\mathcal{M}$  if one refrained from measuring  $\mathcal{M}$ . Imagine that a “measurement bank” implements measurements: The agent hands  $\mathcal{M}$  to a teller. Depending on the memory's state, the teller and agent agree on a fee, which the agent pays with a charged battery. The teller measures  $\mathcal{M}$ ; announces the outcome; resets  $\mathcal{M}$ ; stores the battery's work contents in a vault; and returns the reset memory  $(|0\rangle\langle 0|, \mathbb{1}_d)$  and the empty battery  $(|0\rangle\langle 0|, W|W\rangle\langle W|)$  to the agent.

wherein  $|\sigma| := \sqrt{\sigma^\dagger \sigma}$  for an operator  $\sigma$ . If  $\mathcal{D}(\rho, \tilde{\rho}) \leq \epsilon$ ,  $\rho$  and  $\tilde{\rho}$  are called  $\epsilon$ -close. States  $R = (\rho, H)$  and  $\tilde{R} = (\tilde{\rho}, H)$  are called  $\epsilon$ -close if they have the same Hamiltonian and if their density operators are  $\epsilon$ -close.

The trace distance has the following operational interpretation [39]. Consider a family of POVMs parameterized by  $\gamma$ :  $\mathcal{M}^\gamma = \{M_i^\gamma\}$ . Consider picking a POVM (a value of  $\gamma$ ), preparing many copies of  $\rho$ , and preparing many copies of some approximation  $\tilde{\rho}$ . Imagine measuring each copy of  $\rho$  with  $\mathcal{M}^\gamma$  and measuring each copy of  $\tilde{\rho}$  with  $\mathcal{M}^\gamma$ . A percentage  $\text{Tr}(M_i^\gamma \rho)$  of the  $\rho$  trials will yield outcome  $i$ , as will a percentage  $\text{Tr}(M_i^\gamma \tilde{\rho})$  of the  $\tilde{\rho}$  trials. The difference  $|\text{Tr}(M_i^\gamma \rho) - \text{Tr}(M_i^\gamma \tilde{\rho})|$  between the  $\rho$  percentage and the  $\tilde{\rho}$  percentage will maximize for some  $i = i_{\max}$ . Imagine identifying the greatest difference  $|\text{Tr}(M_{i_{\max}}^\gamma \rho) - \text{Tr}(M_{i_{\max}}^\gamma \tilde{\rho})|$  for each POVM  $\mathcal{M}^\gamma$ . The largest, across the  $\gamma$ -values, of these greatest differences equals the trace distance between  $\rho$  and  $\tilde{\rho}$ :

$$\mathcal{D}_{\text{tr}}(\rho, \tilde{\rho}) = \max_{\gamma} |\text{Tr}(M_{i_{\max}}^\gamma \rho) - \text{Tr}(M_{i_{\max}}^\gamma \tilde{\rho})|. \quad (12)$$

This operational interpretation of the trace distance might facilitate tests of TRT results about  $\epsilon$ -approximations defined in terms of  $\mathcal{D}_{\text{tr}}$ .

As an example, consider checking the prediction that, from an amount  $W_{\text{cost}}^\epsilon(R)$  of work, thermal operations can generate an approximation  $\tilde{R} = (\tilde{\rho}, H)$  to a quasiclassical  $R$  [2]. Suppose we identify a thermal operation  $\mathcal{E}$  expected to transform the work into a  $\tilde{R}$ . Suppose we can perform  $\mathcal{E}$  in a lab. Suppose we have identified the POVM  $\mathcal{M}^\gamma$  that achieves the maximum in Eq. (12).<sup>11</sup> We can test the theorem as follows: implement  $\mathcal{E}$  on each of many copies of the work resource. Measure, with  $\mathcal{M}^\gamma$ , the state produced by each  $\mathcal{E}$  implementation. Note which percentage of the measurements yields outcome  $i$ , for each  $i$ . Prepare many copies of  $R$ . Measure each copy with  $\mathcal{M}^\gamma$ . Note which percentage of the measurements yields outcome  $i$ . Identify the  $i$ -value  $i_{\max}$  for which the  $R$  percentage differs most from the  $\mathcal{E}$  percentage. Confirm that the difference  $\Delta$  between these percentages equals, at most,  $\epsilon$ .<sup>12</sup>

Though blessed with an operational interpretation, the trace distance introduces embezzlement into TRTs [28, 29, 33]. Suppose that thermal operations cannot transform  $R = (\rho, H_R)$  into  $S = (\sigma, H_S)$ :  $R \not\rightarrow S$ .  $R$  might transform into  $S$  catalytically. We call  $C = (\xi, H_C)$  a *catalyst* if  $R \not\rightarrow S$  while some thermal operation can transform the composition of  $R$  and  $C$  into the composition of  $S$  and  $C$ :  $(R + C) \mapsto (S + C)$ . The catalyst resembles an engine that facilitates, but remains unchanged during, the conversion of heat into work.

Realistic engines degrade:

$$(R + C) \rightarrow (S + \tilde{C}). \quad (13)$$

Suppose that the Hamiltonians are completely degenerate:  $H_R = H_S = H_C = 0$ . For every  $R$  and  $S$ , there exist a  $C$  and an arbitrarily similar  $\tilde{C}$  that satisfy (13). The more work the  $R$ -to- $S$  conversion requires, the larger  $C$  must be [the greater the dimension  $\dim(\mathcal{H}_C)$  of the smallest Hilbert space  $\mathcal{H}_C$  on which  $\xi$  can be defined] [28, App. G.2]. The required work is extracted from  $C$ ; this extraction is called *thermal embezzlement*. Embezzlement degrades  $C$  to  $\tilde{C}$ . But if  $\dim(\mathcal{H}_C)$  is large enough, the final catalyst state  $\tilde{\xi}$  remains within trace distance  $\epsilon$  of  $\xi$ . Embezzlement also in the context of nondegenerate Hamiltonians has been studied [28, 33].

Embezzlement contradicts the spirit of the First Law of Thermodynamics: embezzlement outputs an arbitrary amount of work at a barely detectable cost. Hence theorists have called for the elimination of embezzlement from TRTs. Three strategies have been proposed: catalysts' sizes and energies might be bounded;  $\epsilon$  might be defined in terms of catalysts' sizes; or  $\epsilon$  might be defined in terms of a distance measure other than  $\mathcal{D}_{\text{tr}}$ . On the other hand, the challenge of realizing embezzlement may appeal to experimentalists, whose violation of the First Law's spirit would appeal to theorists.

Agents cannot borrow arbitrarily large, or arbitrarily energetic, catalysts in [33]. Suppose that an agent wishes to catalyze a transformation in which all Hilbert spaces' dimensions, and all Hamiltonians' eigenvalues, are bounded. The agent in [33] can borrow only catalysts whose Hilbert spaces are small:  $\dim(\mathcal{H}_C) \leq d_{\text{bound}}$ , for some fixed value  $d_{\text{bound}}$ . Arbitrary amounts of work cannot be embezzled from such catalysts. Suppose that the agent wants a catalyst  $C$  whose spectrum is unbounded (e.g., a harmonic oscillator). Suppose that the partition function  $Z_C := \text{Tr}(e^{-\beta H_C})$  associated with the catalyst's Hamiltonian is finite. The catalysts of this sort that the agent can borrow have bounded average energies:  $\text{Tr}(\xi H_C) \leq E_{\text{bound}}$ . The agent cannot embezzle from these catalysts. In the context of bounded-spectrum catalysts, and in the context of unbounded-spectrum catalysts associated with finite  $Z$ 's, free operations in [33] are designed to preclude embezzlement.

<sup>11</sup> Which  $\mathcal{M}^\gamma$  achieves the maximum follows from the forms of  $R$  and  $\tilde{R}$ . The experimentalist chooses the form of  $R$ . The form of the  $\tilde{R}$  produced by the thermal operation is described in [2, Suppl. Note 4]. (I have assumed that the  $\tilde{R}$  produced by the thermal operation is the state produced in the experiment. But no experiment realizes a theoretical model exactly. This discrepancy should be incorporated into calculations of errors.)

<sup>12</sup> I have ignored limitations on the experimentalist. For example, I have imagined that  $\mathcal{E}$  can be implemented infinitely many times and that infinitely many copies of  $\rho$  can be prepared. They cannot. These limitations should be incorporated into the error associated with  $\Delta$ .

These limitations on catalysts' sizes and energies have pros and cons. The finite- $Z_C$  assumption “holds for all systems for which the canonical ensemble is well-defined” [33]. The assumption seems practical. Yet “there will be specific cases of infinite-dimensional Hamiltonians where simply bounds on average energy do not give explicit bounds on thermal embezzling error” [33]. Examples include the hydrogen atom [71]. The restrictions on infinite-dimensional  $C$ 's do not eradicate embezzling from TRTs. Additionally, bounds on dimension are related to bounds on energy [33], so the two restrictions form a cohesive family. Yet one restriction that eliminated all embezzlement would be more satisfying.

One might eliminate embezzlement from TRTs instead by incorporating catalysts' sizes into the definition of  $\epsilon$ -approximation. In [28, App. F4],  $\tilde{C}$  is called “ $\epsilon$ -close” to  $C$  if

$$\mathcal{D}_{\text{tr}}(\xi, \tilde{\xi}) \leq \frac{\epsilon}{\log(\dim(\mathcal{H}_C))}. \quad (14)$$

This definition reflects the unphysicality of arbitrarily large catalysts. Yet the definition destroys the relevance of one-shot information measures to  $\epsilon$ -approximate catalysis. Suppose that  $\epsilon$ -approximation is defined in terms of the trace distance. Whether  $R$  can be  $\epsilon$ -catalyzed into  $S$  depends on the values of Rényi divergences  $D_\alpha$ . Now, suppose that  $\epsilon$ -catalysis is defined as in Ineq. (14). Whether  $R$  can be  $\epsilon$ -catalyzed into  $S$  depends only on the relative entropy  $D_1$ . Testing one-shot statistical mechanics involves the observation of signatures of  $D_{\alpha \neq 1}$ . Defining  $\epsilon$ -approximate catalysis in terms of catalysts' sizes impedes the testing of one-shot theory.

Third,  $\epsilon$ -approximation could be defined in terms of distance measures other than the trace distance. Many TRT predictions hold if the definition depends on any contractive metric [29, Sec. 6.1]. The *work distance*, for example, has an operational interpretation and has relevance to one-shot entropies [28, App. G.3]. Suppose that some thermal operation maps  $R + C$  to  $\widetilde{S + C}$ . Suppose that, by investing an amount  $\mathcal{D}_{\text{work}}$  of work, one can map  $\widetilde{S + C}$  to  $S + C$  by a catalytic thermal operation.  $\mathcal{D}_{\text{work}}$  is called the *work distance* between  $\widetilde{S + C}$  and  $S + C$ . On the plus side, how much work one can extract from  $R + C \mapsto \widetilde{S + C}$  depends on Rényi divergences  $D_{\alpha \neq 1}$ . On the downside,  $\mathcal{D}_{\text{work}}$  lacks information-theoretic properties of distance measures such as  $\mathcal{D}_{\text{tr}}$ . In terms of which distance measure  $\epsilon$ -approximation should be defined remains undetermined.

I have discussed strategies for eliminating embezzlement from TRTs. Embezzlement appears to merit elimination because it contradicts the spirit of the First Law. The spirit, not the letter. Embezzlement seems physically realizable, in principle. Detecting embezzlement could push experimentalists' abilities to distinguish between close-together states  $C$  and  $\tilde{C}$ . I hope that that challenge, and the chance to contradict the First Law's spirit, attracts experimentalists.

### 5. Can a definition of $\epsilon$ -smoothing more naturally model errors and failure?

The smoothing parameter (alternatively, the error tolerance or failure probability)  $\epsilon$  was introduced in Sec. III B 4. I detailed the operational interpretation for the  $\epsilon$  defined in terms of the trace distance. This operational interpretation involves an optimal POVM  $\mathcal{M}^\gamma$ . By measuring  $\mathcal{M}^\gamma$  in many trials, one might verify whether the  $S$  outputted by an experimental implementation of some thermal operation is  $\epsilon$ -close to the desired state  $R$ . I mentioned shortcomings of this verification scheme, as well as the embezzlement problem. Another problem plagues  $\epsilon$ : Though the “error tolerance”  $\epsilon$  is defined in one-shot information theory, errors seem unable to manifest in single shots of statistical mechanical protocols. Another definition of “error tolerance” might suit TRTs better.

Let us detail how  $\epsilon$  quantifies error in Sec. III B 4. Suppose you wish to measure, with the optimal  $\mathcal{M}^\gamma$ , each of  $n$  copies of  $\rho$ . Unable to prepare  $\rho$  precisely, you create and measure  $\tilde{\rho}$  instead. Your measurements' outcomes differ from the ideal measurements' outcomes in a percentage  $\epsilon$  of the trials, as  $n \rightarrow \infty$ .  $\epsilon$  quantifies the error introduced into your measurement statistics by your substitution of  $\tilde{\rho}$  for  $\rho$ .

The association of  $\epsilon$  with error probability is justified by a limit in which the number  $n$  of trials approaches infinity. No error occurs in any single trial; nor does  $\epsilon$  signify the probability that some trial will suffer an error. Yet  $\epsilon$  is defined in one-shot information theory, which describes finite numbers.<sup>13</sup> This paradox suggests that another definition of  $\epsilon$  might suit TRTs better [49].

How can  $\epsilon$  manifest in single shots? Data compression offers an example. Consider compressing a random variable  $X$  into the smallest possible number  $k$  of bits. Suppose that  $X$  has a probability  $p_i \neq 0$  of assuming the value  $x_i$ , for  $i = 1, 2, \dots, d$ . One bit occupies one of two possible states. Hence a set of  $k$  bits occupies one of  $2^k$  possible states.

<sup>13</sup>Granted, one-shot information theory describes the simultaneous processing of finite numbers of *copies of a probability distribution or quantum state*. The finite numbers referred to above are numbers of *sequential trials*. The TRT manifestation of  $\epsilon$  does not contradict one-shot information theory. Yet the former contradicts the spirit of the latter.

We associate each possible state of  $X$  with one possible state of the bits. Since  $X$  assumes one of  $d$  possible values, the bits must occupy one of  $2^k \geq d$  possible states. We can compress  $X$  into, at fewest,  $k = \lceil \log(d) \rceil$  bits.

If we used fewer bits, then upon decompressing, we would have some probability of failing to recover the value of  $X$ . Suppose that  $X$  probably does not assume the values  $x_1, x_2, \dots, x_m$ . These values' probabilities sum to some tiny number  $\epsilon$ :  $p_1 + p_2 + \dots + p_m = \epsilon$ . We can pretend, via a protocol called *smoothing*, that  $X$  will assume none of these values [34]. We can compress  $X$  almost faithfully into  $k' = \lceil \log(d - m) \rceil$  bits. Decompressing, we have a probability  $\epsilon$  of failing to recover the value of  $X$ , a probability  $\epsilon$  of introducing an error into our representation of  $X$ . Hence the names *failure probability* and *error tolerance*. Smoothing thus has a natural interpretation in one-shot information-processing tasks.

Smoothing, as I argued, seems to lack a natural interpretation in one-shot statistical mechanics. This lack stems partially from resource theorists' having transplanted the definition of smoothing from information theory into TRTs. Not all transplants bloom in their new environs. Yet smoothing appears relevant to statistical mechanics: Smoothing amounts to ignoring improbable events. Improbable events are ignored in statistical mechanics: Broken eggs are assumed never to reassemble, and smoke spread throughout a room is assumed never to recollect in a fireplace. Since smoothing suits statistical mechanics, but the standard definition of  $\epsilon$  seems unsuited to one-shot statistical mechanics, resource theorists might tailor smoothing to TRTs.

## 6. How should TRTs model continuous spectra?

Most TRT predictions concern discrete energies. Yet many real physical systems—quantum environments and classical systems—have continuous spectra. We may need to extend TRT predictions to continuous spectra, to model real systems with TRTs.

In many TRT arguments (e.g., [2, 28–31, 63, 72, 73]), spectra are assumed to be discrete. Simplicity and mathematical convenience justify the assumption. But mathematical convenience trades off with physicality. Resource theorists should check rigorously the limit as the spacing between levels vanishes. TRT predictions have been expected to govern continuous spectra, but limits misbehave.

The need for continuous spectra in TRTs is discussed in [31, 63]. A battery is modeled as a weight that stores free energy as gravitational potential energy. The gravitational-energy operator  $H_{\text{grav}} = mgx$  depends on the position operator  $x$ , whose spectrum is continuous. Skrzypczyk *et al.* argue that, if work is extracted from a finite number  $n$  of bath qubits, the spectrum of  $H_{\text{grav}}$  can be approximated as discrete. One can choose for the spacing between consecutive energies to be  $\mathcal{E} = \epsilon/n^2$ , for some fixed energy  $\epsilon$ . This approximation leads to an error of order  $1/n$  [63, App. G]. In the asymptotic limit as  $n \rightarrow \infty$ , the spacing  $\mathcal{E}$  vanishes. Whether continuous spectra should be approximated as discrete in other TRT contexts merits investigation.

Incorporating continuous spectra into TRTs could facilitate the modeling of classical systems and quantum environments with which TRTs might be realized physically. Many TRT results concern quasiclassical systems, whose density operators  $\rho$  commute with their Hamiltonians  $H$ :  $[\rho, H] = 0$  (e.g., [2, 28–31, 63, 72, 73]). Such quasiclassical results might be realized more easily than quantum results: Classical platforms, such as DNA and colloidal particles, have been used to test small-scale statistical mechanics such as fluctuation relations [6–13]. These classical platforms might be recycled to realize TRTs. Classical testbeds may have continuous spectra [74]. Hence we may need to extend quasiclassical theorems to continuous spectra, or to justify the coarse-graining of real systems' continuous spectra into discrete spectra, to harness existing platforms to realize TRTs.

Like classical systems, many environments of open quantum systems have continuous spectra [75]. TRTs model systems coupled to environments—to heat reservoirs [1]. Insofar as realistic reservoirs have continuous spectra, their TRT analogs should. Some reservoirs are modeled with quantum field theories, discussed in Sec. III C 1, which have continuous spectra. For example, consider a laser mode in a cavity, coupled to the surrounding room by leaky mirrors [75]. This laser is related to the Jaynes-Cummings model, which Åberg introduced into TRTs [60]. The electromagnetic field outside the cavity has a continuum of frequencies. Modeling such QFTs—required to realize TRTs with common systems like lasers—invites us to incorporate continua into TRTs.

## C. More-out-of-the-way opportunities

Physical realizations of TRTs require confrontation of the foregoing nine challenges. The following two challenges appear less crucial. Yet the following could lead to realizations with physical platforms that TRTs could not model with just adjustments discussed above. First, modeling quantum field theories with TRTs could facilitate quantum-optics and condensed-matter realizations. Second, TRTs might reach physical platforms via fluctuation relations. Fluctuation relations describe small scales, as TRTs can, and have withstood experimental tests.

### 1. Can TRTs model quantum field theories?

Quantum field theories (QFTs) represent thermodynamic systems that range from lasers to condensed matter to black holes. Quantum optics and condensed matter are increasingly controllable and conscripted for quantum computation. Holography is shedding the light of quantum information on black holes [76], to which one-shot information theory has been applied [77]. Modeling QFTs with TRTs could unlock testbeds and applications for TRTs. Further motivation appears in [61]: The authors study the “unlocking,” aided by an external field, of work stored in coherence. The field suffers a back reaction. Approximating the field as classical neglects the back reaction. To calculate the unlocking’s cost, one must quantize the field.

I will overview steps with which we can incorporate QFTs into TRTs: an introduction of Fock space into TRTs, more attention to the number operator in TRTs, and more attention to unbounded spectra. Instead of studying general Hamiltonians and qubits, resource theorists would need to focus on quantum-optics and condensed-matter Hamiltonians. Groundwork for these expansions has been laid in [30, 31, 60].

Specifying a state in a TRT involves specifying a Hilbert space. Quantum states in QFTs are defined on Fock spaces. Since Fock spaces are Hilbert spaces, TRTs offer hope for modeling QFTs. A popular Fock-space basis consists of eigenstates  $|n\rangle$  of the particle-number operator  $N$ . A similar operator was introduced into TRTs in [30, 31]. The number-like operators in [30, 31] have bounded spectra, whereas arbitrarily many particles can populate a QFT. Spectra were bounded in keeping with the boundedness of the energy spectra in many TRT proofs. TRT spectra are bounded “for convenience” or “for simplicity.” This phrasing suggests that proofs are believed to extend easily to unbounded spectra. That the proofs extend merits checking, and how best to model unbounded spectra in TRTs merits consideration. Continuous spectra, their relevance to QFTs, and their incarnation in TRTs is discussed in Sec. IIIB 6.

Unbounded spectra and lasers forayed into TRTs in [33, 60, 63]. In [33], catalysts have unbounded spectra. The expectation value of each catalyst’s Hamiltonian remains below some finite value  $E$ . The effective cutoff reduces the problem to a finite-spectrum problem. The cutoff is justified with the finiteness of real systems’ energies.

Cutoffs feature also in [63]: a weight (such as a stone) that has gravitational potential energy models a battery. The Hamiltonian  $H_{\text{grav}} = mgx$  governs the weight, wherein the position operator  $x$  denotes the weight’s height. The position operator has an unbounded spectrum. Yet the battery’s height must lie below some cutoff. Once a weight reaches a certain height above the Earth,  $mgx$  approximates the stone’s potential energy poorly. Calculating a limit as the cutoff approaches infinity might incorporate truly unbounded spectra fundamentally into TRTs.

In [60], Åberg treats a doubly infinite ladder and a half-infinite ladder as environments. The doubly infinite ladder’s energy spectrum runs from  $-\infty$  to  $\infty$ . Though unphysical, the ladder simplifies proofs. Åberg extends these proofs to the more physical half-infinite ladder, whose energy spectrum is bounded from below but not from above. This harmonic oscillator could pave the TRT path toward QFTs, which consist of oscillators that vibrate at various frequencies. One workhorse of QFT is the Jaynes-Cummings model of quantum optics. The model describes how matter exchanges energy with an electromagnetic field. Åberg models a Jaynes-Cummings-like system with his framework. “[T]o what extent such a generalized Jaynes-Cummings interaction can be obtained, or at least approximated, within realistic systems,” he writes, “is left as an open question” [60, App. E.2].

Åberg discusses also lasers, a stalwart of many labs. Laser light is often modeled with a coherent state,

$$|\alpha\rangle := e^{-\frac{1}{2}|\alpha|^2} \sum_{l=0}^{\infty} \frac{\alpha^l}{\sqrt{l!}} |l\rangle, \quad (15)$$

wherein  $\{|l\rangle\}$  denotes the energy eigenbasis and  $|\alpha\rangle$  denotes an eigenstate of the annihilation operator  $a$ :  $a|\alpha\rangle = \alpha|\alpha\rangle$  [75]. Åberg studies, rather than coherent states  $|\alpha\rangle$ , uniform superpositions  $|\eta_{L,l_0}\rangle := \sum_{l=0}^{L-1} \frac{1}{\sqrt{L}} |l_0 + l\rangle$  of neighboring energy eigenstates. Extending his analysis to the  $|\alpha\rangle$  can shift TRTs toward modeling lasers and other real systems described by QFTs.

### 2. Can fluctuation theorems bridge TRTs to experiments?

*Fluctuation theorems* are statistical mechanical predictions about systems arbitrarily far from equilibrium. Fluctuation theorems have withstood experimental tests and describe small scales. Can TRTs reach experiments via fluctuation theorems? If so, how? I will survey progress toward answers.

Fluctuation theorems describe the deviations, from equilibrium values, of outcomes of measurements of statistical mechanical systems. Consider a system coupled to a bath at inverse temperature  $\beta$ . Suppose that a perturbation changes the system’s Hamiltonian from  $H_i$  to  $H_f$ . If the system consists of a gas, it might undergo compression [78].



If the system consist of a trapped ion, a laser might induce a time-evolving field [13]. Consider time-reversing the perturbation. Statistics that characterize the forward process can be related to statistics that characterize the reverse process, to differences  $\Delta F$  between free energies, and to deviations from equilibrium statistics. Such relations are fluctuation theorems.

Examples include Crooks' Theorem and Jarzynski's Equality. Let  $P_{\text{fwd}}(W)$  denote the probability that some forward trial will require an amount  $W$  of work (e.g., to compress the gas). Let  $P_{\text{rev}}(-W)$  denote the probability that some reverse trial will output an amount  $W$ . According to *Crooks' Theorem* [78],

$$\frac{P_{\text{fwd}}(W)}{P_{\text{rev}}(-W)} = e^{\beta(W-\Delta F)}. \quad (16)$$

This  $\Delta F = F_f - F_i$  denotes the difference between the free energy  $F_f = -k_B T \ln(Z_f)$  of the equilibrium state  $e^{-\beta H_f}/Z_f$  relative to the final Hamiltonian and the free energy  $F_i$  of the equilibrium state relative to  $H_i$ . Whereas  $\Delta F$  characterizes equilibrium states,  $W$  characterizes a nonequilibrium process. Jarzynski's Equality informs us about nonequilibrium, difficult to describe theoretically, via the equilibrium  $\Delta F$ , easier to describe theoretically. Conversely, from data about nonequilibrium trials, easier to realize in practice, we can infer the value of the equilibrium  $\Delta F$ , an unrealizable ideal. Multiplying each side of Crooks' Theorem by  $P_{\text{rev}}(-W)e^{-\beta W}$ , then integrating over  $W$ , yields *Jarzynski's Equality* [79]:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \quad (17)$$

Fluctuation theorems were derived first from classical mechanics. They have been extended to quantum systems [80] and generalized with information theory (e.g., [81]). Experimental tests have involved DNA [6–8], trapped colloidal particles [11], single-electron boxes [12], trapped-ion harmonic oscillators [13], and other platforms.

Fluctuation theorems and TRTs describe similar (and, in some cases, the same) problems. First, both frameworks describe arbitrarily-far-from equilibrium processes. Second, each framework features work, entropy, and new derivations of the Second Law. Third, both frameworks describe small scales. I discussed the relationship between small scales and TRTs in Sec. III A 1. The systems that obey fluctuation theorems most noticeably are small: Deviations from equilibrium behaviors decay with system size [38]. Hence we can most easily detect deviations in small systems, such as single strands of DNA. Fourth, heat exchanges governed by Crooks' Theorem can be modeled with TRT thermal operations [4]. Since the physics described by fluctuation theorems overlaps with the physics described by TRTs, and since experimentalists have tested fluctuation theorems, fluctuation theorems might bridge TRTs to experiments.

Construction of the bridge has begun but needs expansion. Åberg derived a Crooks-type theorem from one-shot statistical mechanics [3, Suppl. Note 10B]. How to model, with TRTs, a process governed by Crooks' Theorem was detailed in [4]. A fluctuation relation was derived from TRT principles in [5]. In [82], one-shot analogs of asymptotic fluctuation-theorem work quantities were derived. These first steps demonstrate the consistency between fluctuation theorems and TRTs (or one-shot statistical mechanics). How to wield this consistency, to link TRTs to experiments via fluctuation theorems, remains undetermined.

#### IV. CONCLUSIONS

During the past few years, the literature about thermodynamic resource theories has exploded. Loads of lemmas and reams of theorems have been proven. To what extent do they describe physical reality? Now that TRTs have matured, they merit experimental probing. I have presented eleven opportunities in physically realizing TRTs. The challenges range from philosophical to practical, from speculative to expectedly straightforward. These opportunities might generalize to physical realizations of other resource theories, such as the resource theory for coherence [22].

I concentrated mostly on gaps in resource theories, on how theorists might nudge their work toward physical realizations. Yet I hope that the discussion will appeal to experimentalists. An experimentalist opened the Q&A of my seminar two years ago. His colleagues and thermodynamic resource theorists have an unprecedented opportunity to inform each other. Let the informing begin... preferably with more cooperation and charity than during the Q&A.

#### ACKNOWLEDGEMENTS

I am grateful to Fernando Brandão, Lídia del Rio, Ian Durham, Manuel Endres, Tobias Fritz, Alexey Gorshkov, Christopher Jarzynski, David Jennings, Matteo Lostaglio, Evgeny Mozgunov, Varun Narasimhachar, Nelly Ng, John Preskill, Renato Renner, Dean Rickles, Jim Slinkman, Stephanie Wehner, and Mischa Woods for conversations and feedback. This research was supported by an IQIM Fellowship, NSF grant PHY-0803371, and a Virginia

Giloon Fellowship. The Institute for Quantum Information and Matter (IQIM) is an NSF Physics Frontiers Center supported by the Gordon and Betty Moore Foundation. Stephanie Wehner and QuTech offered hospitality at TU Delft during the preparation of this manuscript. Finally, I thank that seminar participant for galvanizing this exploration.

## REFERENCES

- [1] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *Physical Review Letters* **111**, 250404 (2013).
- [2] M. Horodecki and J. Oppenheim, *Nat. Commun.* **4**, 1 (2013).
- [3] J. Åberg, *Nat. Commun.* **4**, 1925 (2013).
- [4] N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, *ArXiv e-prints* (2014), 1409.3878.
- [5] S. Salek and K. Wiesner, *ArXiv e-prints* (2015), 1504.05111.
- [6] A. Alemany and F. Ritort, *Europhysics News* **41**, 27 (2010).
- [7] A. Mossa, M. Manosas, N. Forns, J. M. Huguët, and F. Ritort, *Journal of Statistical Mechanics: Theory and Experiment* **2009**, P02060 (2009).
- [8] M. Manosas, A. Mossa, N. Forns, J. Huguët, and F. Ritort, *Journal of Statistical Mechanics: Theory and Experiment* **2**, 61 (2009), 0902.3634.
- [9] Bérut, A., Petrosyan, A., and Ciliberto, S., *EPL* **103**, 60002 (2013).
- [10] Y. Jun, M. Gavrilov, and J. Bechhoefer, *Phys. Rev. Lett.* **113**, 190601 (2014).
- [11] V. Blicke, T. Speck, L. Helden, U. Seifert, and C. Bechinger, *Phys. Rev. Lett.* **96**, 070603 (2006).
- [12] O.-P. Saira *et al.*, *Phys. Rev. Lett.* **109**, 180601 (2012).
- [13] S. An *et al.*, *Nature Physics* **11**, 193 (2015), 1409.4485.
- [14] A. S. Eddington, *The Nature of the Physical World* (MacMillan, 1929).
- [15] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, *ArXiv e-prints* (2015), 1505.07835.
- [16] B. Coecke, T. Fritz, and R. W. Spekkens, *ArXiv e-prints* (2014), 1409.5531.
- [17] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [18] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, *Rev. Mod. Phys.* **79**, 555 (2007).
- [19] I. Marvian and R. W. Spekkens, *New Journal of Physics* **15**, 033001 (2013).
- [20] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner, *New J. Phys.* **8**, 58 (2006).
- [21] V. Veitch, S. A. Hamed Mousavian, D. Gottesman, and J. Emerson, *New Journal of Physics* **16**, 013009 (2014), 1307.7171.
- [22] A. Winter and D. Yang, *ArXiv e-prints* (2015), 1506.07975.
- [23] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, *Science (New York, N.Y.)* **296**, 1832 (2002).
- [24] C. Bustamante, J. Liphardt, and F. Ritort, *Physics Today* **58**, 43 (2005), cond-mat/0511629.
- [25] L. P. Faucheux, L. S. Bourdieu, P. D. Kaplan, and A. J. Libchaber, *Phys. Rev. Lett.* **74**, 1504 (1995).
- [26] J. Cheng *et al.*, *Phys. Rev. Lett.* **109**, 238104 (2012).
- [27] J. P. Pekola, *Nature Physics* **11**, 118123 (2015).
- [28] F. G. S. L. Brandão, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner, *Proc. Nat. Acad. Sci. USA* **112**, 3275 (2015).
- [29] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, *Physics Reports* **583**, 1 (2015), The resource theory of informational nonequilibrium in thermodynamics.
- [30] N. Yunger Halpern and J. M. Renes, *ArXiv e-prints* (2014), 1409.3998.
- [31] N. Yunger Halpern, *ArXiv e-prints* (2014), 1409.7845.
- [32] M. Lostaglio, D. Jennings, and T. Rudolph, *Nature Communications* **6**, 6383 (2015), 1405.2188.
- [33] N. H. Y. Ng, L. Mančinská, C. Cirstoiu, J. Eisert, and S. Wehner, *ArXiv e-prints* (2014), 1405.3039.
- [34] R. Renner, *Security of Quantum Key Distribution*, PhD thesis, PhD Thesis, 2005, 2005.
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
- [36] S. Vinjanampathy and J. Anders, *ArXiv e-prints* (2015), 1508.06099.
- [37] C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
- [38] D. V. Schroeder, *An introduction to thermal physics* (Addison Wesley, San Francisco, CA, 2000).
- [39] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
- [40] A. Rényi, On measures of entropy and information, in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, 1961.
- [41] T. van Erven and P. Harremoës, *ArXiv e-prints* (2012), 1206.2459.
- [42] D. Petz, *Reports on Mathematical Physics* **23**, 57 (1986).
- [43] N. Datta, *IEEE T. Inform. Theory* **55**, 2816 (2009).
- [44] M. Tomamichel, *arXiv:1203.2142* (2012).
- [45] F. Buscemi and N. Datta, *ArXiv e-prints* (2009), 0902.0158.
- [46] F. G. S. L. Brandao and N. Datta, *ArXiv e-prints* (2009), 0905.2673.
- [47] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, *Physical Review X* **5**, 021001 (2015), 1410.4572.
- [48] L. Mazza, D. Rossini, R. Fazio, and M. Endres, *New Journal of Physics* **17**, 013015 (2015), 1408.4672.
- [49] R. Renner, private communication, 2015.
- [50] C. M. Caves, *Phys. Rev. D* **26**, 1817 (1982).

- [51] J. P. Joule, *Philosophical Magazine* **xxvii** (1845).
- [52] D. Reeb and M. M. Wolf, *New Journal of Physics* **16**, 103011 (2014), 1306.4352.
- [53] H. Tajima and M. Hayashi, *ArXiv e-prints* (2014), 1405.6457.
- [54] M. P. Woods, N. Ng, and S. Wehner, *ArXiv e-prints* (2015), 1506.02322.
- [55] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, *Proc. Nat. Acad. Sci. USA* **111**, 13786 (2014).
- [56] A. S. L. Malabarba, A. J. Short, and P. Kammerlander, *New Journal of Physics* **17**, 045027 (2015).
- [57] D. Gottesman, *ArXiv e-prints* (2009), 0904.2557.
- [58] A. D. Córcoles *et al.*, *Nature Communications* **6**, 6979 (2015), 1410.6419.
- [59] M. Horodecki *et al.*, *Phys. Rev. Lett.* **90**, 100402 (2003).
- [60] J. Åberg, *Physical Review Letters* **113**, 150402 (2014), 1304.1060.
- [61] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, *ArXiv e-prints* (2015), 1506.07875.
- [62] D. Jennings, private communication, 2015.
- [63] P. Skrzypczyk, A. J. Short, and S. Popescu, *arXiv:1302.2811* (2013).
- [64] H. Wilming, R. Gallego, and J. Eisert, *ArXiv e-prints* (2014), 1411.3754.
- [65] Á. M. Alhambra, J. Oppenheim, and C. Perry, *ArXiv e-prints* (2015), 1504.00020.
- [66] L. Szilard, *Zeitschrift für Physik* **53**, 840 (1929).
- [67] R. Landauer, *IBM J. Res. Dev.* **5**, 183 (1961).
- [68] M. Lostaglio, private communication, 2015.
- [69] M. Navascués and L. P. García-Pintos, *ArXiv e-prints* (2015), 1501.02597.
- [70] J. Oppenheim, In “open-problem session”, 2015.
- [71] N. Ng, private communication, 2015.
- [72] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, *Int. J. Theor. Phys.* **39**, 2717 (2000).
- [73] M. Horodecki, P. Horodecki, and J. Oppenheim, *Phys. Rev. A* **67**, 062104 (2003).
- [74] C. Jarzynski, H. T. Quan, and S. Rahav, *ArXiv e-prints* (2015), 1507.05763.
- [75] G. S. Agarwal, *Quantum Optics* (Cambridge U. P., 2013).
- [76] D. Harlow, *ArXiv e-prints* (2014), 1409.1231.
- [77] B. Czech, P. Hayden, N. Lashkari, and B. Swingle, *ArXiv e-prints* (2014), 1410.1540.
- [78] G. E. Crooks, *Physical Review E* **60**, 2721 (1999).
- [79] C. Jarzynski, *Physical Review Letters* **78**, 2690 (1997).
- [80] M. Campisi, P. Hänggi, and P. Talkner, *Rev. Mod. Phys.* **83**, 771 (2011).
- [81] T. Sagawa and M. Ueda, *Physical Review Letters* **104**, 090602 (2010), 0907.4914.
- [82] N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, *ArXiv e-prints* (2015), 1505.06217.